On randomization, modeling and experimental design; a new example in an old discussion

Vos, A.F.

1987

SERIE RESEARCH MEMORANDA

ON RANDOMIZATION, MODELING AND EXPERIMENTAL DESIGN; A NEW EXAMPLE IN AN OLD DISCUSSION

Aart F. de Vos

Researchmemorandum 1987-53 Dec. '87

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM

On Randomization, modeling and experimental design; a new example in an old discussion.

Aart F. de Vos *)

Abstract

This paper is a modern variant of one of the early debates in statistics: the question whether to randomize experiments or to rely on model building. For historical reasons the -very general- discussion is in terms of a famous problem: the estimation of the difference in yield of two breeds of corn. In a simplified setup it is shown that the use of a simple robust model to take variations in soil fertility into account leads to inference and experimental design that outperforms the inference from randomized experiments by far. The subject is tackled from an econometric viewpoint, supplemented with some references to old and new statistical literature.

October 1987

---

*)Faculteit der Economische Wetenschappen en Econometrie
Vakgroep Econometrie
Vrije Universiteit
Postbus 7161
1007 MC Amsterdam

On Randomization, modeling and experimental design; a new example in an old discussion.

Aart F. de Vos; October 1987.

This paper is a modern variant of one of the early debates in statistics: the question whether to randomize experiments or to rely on model building. For historical reasons the -very general- discussion is in terms of a famous problem: the estimation of the difference in yield of two breeds of corn. It is shown that the use of a simple robust model to take variations in soil fertility into account leads to inference and experimental design that outperforms the inference from randomized experiments by far. The point is not new, but the example is simple and powerful. As for most statisticians randomization is still a paradigm such an example may contribute to the renewed discussion. The subject is tackled from an econometric viewpoint, supplemented with some references to old and new statistical literature.

My distrust of randomization

Personal intuitive distrust of randomization is the main motive behind this paper. I even think that this was the first distrust of statistical methodology that occurred to me. About twenty years ago, during my first courses in Statistics, I heard about the historical discussions on experimental design in Statistics. The quintessence, written down from my memory (second thoughts after reading some literature will follow) was as follows:

Suppose one has a piece of land subdivided into a large number of plots. Furthermore, one has two breeds of corn and one wants to know which has a higher yield. What is the best experimental design, consisting of selecting plots to seed both breeds, measuring yields per plot and drawing conclusions?

The problem is that there are variations in soil fertility: neighboring plots may be supposed to have correlated deviations from mean fertility. Therefore it seems unwise to use all plots in the south for breed 1 and

all plots in the north for breed 2: differences in yield may be caused by differences in fertility. One of the alternatives with more intuitive appeal is to consider the land as a chessboard, using the white plots for breed 1 and the black plots for breed 2. Many devices like that were suggested in the early literature on experimental design. But then the father of modern statistics, Sir Ronald Fisher arrived and said that this procedure would lead to biased judgments because the differences in fertility faced by the two breeds would have smaller variances than the differences within each group, so estimation and testing procedures that neglect this fact would be biased. So would other designs yield other biases. He argued that there is only one satisfactory solution: randomizing the experiment. This enables the statistician to make mathematically justified statements about the probability of outcomes of the experiment (before drawing the sample).
My uneasy feeling about this was primarily based on what I now know to be the problem of "unwanted random assignments", first put forward by "Student" (1937): suppose one performs the experiment and by coincidence the randomized sample is just the chessboard, or, even worse, the north-south division. The outcome of the experiment is then equal to something one wanted to avoid, how can such a procedure be justified?
Being an econometrician I did not worry much about the problem: econometricians tend to work with the horrible experimental designs of administrators and other people who are not interested in inference at all, or -in time series- with experimental designs that must be attributed to Keynes' "animal spirits". They are supposed to make the best of it. Perhaps this is the reason that mainstream econometrics is concerned with conditional models: at least these have the advantage that randomization of the conditioning variables is not essential (Leamer(1983)).

The old doubts got my renewed interest when I was recently confronted with a field where randomization plays an even more curious part. In auditing samples are drawn from items in financial statements. The items may wholly or partially be wrong. The goal is to make statements on the total error amount in the population, based upon known errors in a sample. The predominant methodology is "monetary unit sampling" (MUS):

pretend that all items are put in random order and that each n'th dollar is drawn. Check the items containing drawn dollars and compute the number of false dollars, pretending that in a partially wrong item the false dollars are all in the beginning (or at the end, but make a choice before drawing). In this example a more serious aspect of randomization becomes apparent: deliberately a lot of information is put aside to obtain "objective" statements.

Some thoughts on randomization are further worked out in the following sections. First general aspects of the modeling alternative to randomization are discussed. Then some literature is surveyed, the classic discussion as well as some recent contributions. Next the central example is worked out, showing that in the corn breeding case the use of a simple robust model to take variations in soil fertility into account leads to inference and experimental design that outperforms the inference from randomized experiments by far. Finally some remarks are made on the auditing problem, where the problems with modeling appear to be more complex.

Natural and artificial randomization.

The concept of randomness in the dependent variable is in econometric model building seldom connected with active randomization. The econometric axiom seems to be that residuals that remain after the process of modeling and diagnostic testing are random. In fact a considerable part of modern econometric activity, especially in the field of time series, may be seen as searching the "natural" random component in the dependent variable. What remains after appropriate modeling is random by definition.
Whether this definition of randomness must be mistrusted depends on the beliefs one has in the human capacity to remove non-randomness. Orthodox classical statisticians apparently do not believe in model building: they replace correlation structures that may be very useful in model-based inference by artificial randomness. To use a variant on a famous image: to study eggs they make omelets.

On the other hand no model builder believes his residuals to be truly random. Unfortunately there is by definition no way to prove that something is random. In her lucid treatment of randomness Lopes (1982) states cursive: "*no one at all* has ever seen what randomness is like" (p 630). Distrust against the model builder's definition of randomness may only be justified by proving some dependence in his residuals, a process so close to model building that it seems an inappropriate way to condemn it. Perhaps this is the reason why artificial randomness is often used as an absolute device.

The discussions in some old and new literature.

A nice description of the historical discussions on randomization is given by Picard (1980). My summary is mainly based on this. Fisher exposed his ideas on the necessity of randomization starting first in Fisher (1925). He rejected all previous and subsequent systematic designs among which the "Knight move, the "Diagonal"and the "Half Strip Drill". His main argument against these proposals was that these gave no valid estimate of the error component. He admitted that real errors were probably smaller than in the randomized experiment, but argued that their computation was unclear.
His main opponent was "Student". The essence of the debate is in "Student"'s reply on Fisher discussing "Student"(1936): "He says to me 'your half drill strips have no validity and conclusions cannot be drawn from them'. I say to him 'your errors are so large that no conclusions are drawn'. Neither of these criticisms is true and one is about as good as any other." The end of this discussion is the posthumous article "Student" (1937), introducing the problem of bad random designs, proposing another design, the "Double Sandwich".
Though much has been written on experimental design since, many questions concerning the relative merits of systematic design as opposed to randomization have yet to be resolved (Picard (1980, p 57) and unwanted random assignments are still seen as a "philosophical problem" (Holschuh (1980 p.44)).

This philosophical problem touches a controversy on statistical inference that is fundamental. One may question whether the fact that an experiment belongs to a class of experiments (samples) that *on the average* lead to proper inference is a good ground to justify inference based on a unique experiment. The irrelevance of sampling properties is essentially the basic argument for Bayesian statistics, described by Box and Tiao (1973, p.72) as:

"In sampling theory (...) the probabilities we calculate refer to the frequency with which different values of statistics (arising from sets of data *other* than those which have actually happened) could occur for some *fixed but unknown values of the parameters*. (....)
By contrast, in Bayesian analysis, inferences are based on probabilities associated with *different* values of parameters which could have given rise to the *fixed* set of data which has actually occurred."

In other words: once the sample has been drawn inference only depends on this sample. And if the selection method does not involve the parameters one wants to say something about, the way the sample is drawn is irrelevant. This is not only a Bayesian viewpoint, but a general feature of model based inference. The likelihood may involve the postulated natural random component as well as the sampling scheme, and may be decomposed into the likelihood of the sample and the likelihood of the endogenous variable given the sample. Thus the sampling scheme need not be random, it may be chosen to be most informative.
The main argument for randomization that remains in this view is that the sample must be representative for the population one wants to say something about. This argument seems more relevant in survey sampling than in experimental design. Fienberg and Tanur (1987) give a review of the literature on survey sampling and experimental design, their differences and resemblances. Striking is the multitude of methodologies that appear in this review, for a large part differing in their appeal to randomization arguments. Orthodox pure randomizers appear to be rare, but mainstream statistical literature applies forms of restricted randomization. The model building view is slowly gaining, which is

illustrated by the fact that Fienberg and Tanur end their article with a fundamental treatment of issues in model based inference.
An interesting vision trying to reconcile randomizers and model builders is represented by Cheng and Li (1987). They put the matter as a choice of criterion. Randomizers are minimax: their action is the best if circumstances are the worst, the suggested models being completely wrong. Model builders minimize the Mean Squared Error, conditional upon their model. Application of a generalized criterion, ranging from minimax to MSE leads to some solution in between.

Before turning to my down-to-earth example let me paraphrase the discussion:
The funny thing about randomization is that as soon as one knows why it is done, a better alternative is available. In some cases it will be possible to incorporate knowledge on the disturbing factor into a model. Investigating the effect of a drug while blood pressure may be supposed to effect it asks for a model using blood pressure as an explanatory variable. In other cases -like soil fertility- there is no explicit measure of the disturbing factor but it is known which observations may be supposed to suffer from similar biases. In other words, something is known about the correlation structure of the disturbances. This is the case I will address.

A simple problem: experimental design in one dimension.

As an example I take the plots of land from the prewar discussion, but to simplify things I suppose that they are all in one line and equidistant. The variation in soil fertility then resembles a time series and we may employ the powerful class of ARIMA models to describe correlation structures. The problem looks like:

$$y_i = X\alpha + u_i \qquad i=1\ldots n \qquad (1)$$

where y_i is the yield per plot, X a nx2 matrix containing the experimental design: x(i,k)=1 if plot i contains breed k (i=1,2), else 0. α is the 2x1 vector of the yields of both types of corn, we are interested in $\alpha_1-\alpha_2$. Finally u_i follows an ARIMA process. The simplest example is the AR(1) model

$$u_i = \rho u_{i-1} + \epsilon_i \tag{2}$$

with ϵ white noise. This model seems for most situations an appropriate and robust approximation of existent correlation structures. ρ may be estimated, so loosely speaking the mean correlation between neighboring plots is taken into account, and the main assumption is that fertility variations are stationary.

The most relevant aim is an estimate of $\alpha_1-\alpha_2$ with minimal mean squared error: one is not so much interested in the average yields as such on the plots, but assumes implicitly that the difference in yields is more or less stable in other circumstances. I will consider four strategies.

A. Randomization and abstinence from a model

A standard classical approach would be to randomize the experiment. Let us say by tossing a coin for each plot to determine which breed will be used. Next the average yields for the two breeds are compared, using an estimate of σ_u. The autocorrelation is not incorporated, using the argument that the randomized disturbances no longer show correlation. Then the estimate a_1-a_2 of $\alpha_1-\alpha_2$ is unbiased and has variance

$$(n_1^{-1} + n_2^{-1})\ \sigma_u^2$$

and the first thing to note is that it is wise to take $n_1=n_2=n/2$. Restricted randomization may accomplish this leading to:

$$VAR(a_1-a_2)_{rand} = (4/n)\sigma_u^2 = (4/n)\sigma_\epsilon^2 * (1-\rho^2)^{-1} \tag{3}$$

We will come back upon this evaluation of the variance in the treatment of strategy D.

B. Randomization and estimation with a model

A more efficient way to estimate $\alpha_1-\alpha_2$,, incorporating knowledge of ρ, is generalized least squares. This optimal estimation procedure, may, deleting some information from the first plot be performed by ordinary least squares on

$$y_i - \rho y_{i-1} = (x_{1i} - \rho x_{1,i-1})\alpha_1 + (x_{2i} - \rho x_{2,i-1})\alpha_2 + \epsilon_i \qquad (4)$$

The resulting covariance matrix for a_1, a_2 is easily evaluated using the fact that it depends on the occurrence of combinations of both breeds on neighboring plots, so with ρ known it depends on four statistics, or rather three as the total sample size is known. Counting the combinations one gets:

n_{11} times 1,1
n_{21} times 0,1
n_{12} times 1,0
n_{22} times 0,0

the covariance matrix of (a_1,a_2) is the inverse of:

$$\begin{bmatrix} n_{11}(1-\rho)^2 + n_{21}+n_{12}\ \rho^2 & -\rho(n_{21}+n_{12}) \\ -\rho(n_{21}+n_{12}) & n_{22}(1-\rho)^2 + n_{12}+n_{21})\rho^2 \end{bmatrix} = \begin{bmatrix} a & -c \\ -c & b \end{bmatrix}$$

and the variance of a_1-a_2 is

$$[1 \quad -1] \begin{bmatrix} a & -c \\ -c & b \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{a+b-2c}{ab - c^2} \qquad (5)$$

$a + b - 2c$ appears to be $(n_{11}+n_{21}+n_{12}+n_{22})(1-\rho)^2$, thus independent from the design, but $ab-c^2$ is not.
To simplify things I only consider designs where $n_{21}=n_{12}$ and $n_{11}=n_{22}$, known from restricted randomization (Latin Squares).
Some calculations lead to:

$$ab-2c = n_{11}^2(1-\rho)^4 + 2n_{11}n_{12}(1-\rho)^2(1+\rho^2) + n_{12}^2(1-\rho^2)^2$$

which can be rewritten as

$$ab\text{-}2c = (n_{11} + n_{12})((1-\rho)^4 n_{11} + (1-\rho^2)^2\, n_{12}) \tag{6}$$

In the randomized experiment the expected outcome is $n_{11}=n_{21}=n_{12}=n_{22}=n/4$. Conditioning upon this (which is almost the same as taking expectations) the "randomized GLS estimator" (rGLS) has variance:

$$\mathrm{VAR}(a_1 - a_2)_{rGLS} = (4/n)\sigma_{\epsilon}{}^{2}{}_{*}\,(1+\rho^2)^{-1} \tag{7}$$

which is smaller than (3) as it should be.

C.Experimental design using a model

The most interesting situation arises when one wonders whether the fact that there is autocorrelation may be used for a better experimental design. This appears to be simple in this case. One has to maximize ab-2c from (6). As

$$(1-\rho^2)^2 - (1-\rho)^4 = 4\rho(1-\rho)^2 > 0 \tag{8}$$

one must (for $\rho>0$) simply take n_{12} as large as possible:
$n_{12} = n/2$, $n_{11} = 0$.

Conclusion: for any value of $\rho>0$ the alternating design is optimal in combination with the GLS estimation procedure. For $\rho = 0$ the design does not matter. For $\rho = 1$ the formulae must be revised somewhat, but it is trivial that one must take the alternating design: subsequent plots with the same breed give no information at all. For $\rho<0$ (8) shows that one has to take as many neighboring plots with the same breed as possible. This case does not seem very relevant in practice.
Confining ourselves to $\rho>0$ the result of the optimal combination of experimental design and estimation procedure appears to be:

$$\mathrm{VAR}(a_1 - a_2)_{opt} = (4/n)\sigma_{\epsilon}{}^{2}{}_{*}\,(1+\rho)^{-2} \tag{9}$$

Table 1 summarizes the efficiency of the estimators (7) and (9) compared to (3)

Table 1 relative efficiency of rGLS and optimal estimators.
(VAR of rand divided by VAR of other estimators)

ρ	rGLS	opt
0	1	1
.1	1.02	1.22
.2	1.08	1.50
.3	1.20	1.86
.4	1.38	2.33
.5	1.67	3.00
.6	2.13	4.00
.7	2.92	5.67
.8	4.56	9.00
.9	9.53	19.00

The possible gains are quite spectacular. Of course one has to keep in mind that in practice ρ is not known and the appropriateness of the AR(1) model may be questioned. This leads to some loss of efficiency. The estimation problem is asymptotically negligible but the effect of the specification problem remains a snag. The robustness of inference within the class of ARMA models with respect to the choice of the model is the main argument to attach little weight to this argument. As Sneek(1985) has shown many members of the ARMA family are very much alike.

A striking fact is that the optimal design does not depend on ρ (supposing $\rho>0$). This suggests that some rather general rule may be given, possibly that alternating designs (in space as well as on a line) are always useful in case of positive correlations.

D.Model-based evaluation of the standard estimator

The difference between mean yields per breed, which I denote as b_1-b_2 will normally be the standard reported estimate of α_1-α_2. Therefore it

is interesting to look at the properties of this estimator under different designs and models.
In the randomized design the variance *conditional upon the outcome of the randomization* - restricted here to the case $n_1=n_2=n/2$ - may be evaluated as follows:

$$b_1-b_2 = \alpha_1-\alpha_2 + (2/n)\ \Sigma_i(s_i u_i)$$

with $s_1 = 1$ if breed 1 is at plot i and $s_1 = -1$ if breed 2 is at plot i.
As $\text{cov}(u_j, u_{j-k}) = \rho^k\ \sigma_u^{\ 2}$

$$\text{Var}(b_1-b_2) = (2/n)^2\sigma_u^{\ 2}(n + \Sigma_{i=1}^{n-1}\ 2\rho^i(m_{is}-m_{id}) \qquad (10)$$

with m_{is} the number of times that plots i places apart contain the same breed and m_{id} the number of times this is different.
The expected value of (10) in the class of randomized results is (3): $E(m_{is}-m_{id}) = 0$. Given the sample this is not true and one might say that (3) is biased.

A rather funny result emerges if one calculates (10) for the alternating design

$$\text{Var}(b_1-b_2)_{alt} = (2/n)^2\sigma_u^{\ 2}(n + \Sigma_{i=1}^{n-1}\ 2\rho^i(-1)^i(n-i)) \qquad (11)$$

asymptotically this is equal to $(2/n)^2\sigma_u^{\ 2}(1-\rho)(1+\rho)^{-1}$, which is the same as that of the optimal estimator in (9). On closer inspection this is as it should be: the classical and the GLS estimators coincide in the alternating design (a proof is given at the and of this paper). This result is probably specific for the AR(1) model. Still it is amusing to conclude that randomizing throws away the optimal design and the optimal estimation procedure, while the optimal design and the classical estimation procedure do the job perfectly, if only a model is used to evaluate the variances.

Non stationary situations

Working conditional upon a model has one drawback: one is not certain that the model is correct. What one would like to show is that this does not matter much: if the model is approximately correct (by estimation and testing this may be assured) the inference should be approximately correct.

In time series the most interesting problems with respect to models that are empirically close but may lead to important differences in inference arise when one model is stationary and one is not. It will be shown that in our example this problem is not severe. And what is more: model based inference using the alternating design is robust, whereas the randomization concept breaks down when non-stationarity is involved.

An interesting feature of the formulae (7) and (9) is that the results remain valid for ρ=>1. ρ=1, the "random walk in fertility" may seem an odd choice but nonstationary changes in fertility are certainly a possibility, if only because in different places in the world there are widely different circumstances. Models that result as the sum of some slow nonstationary change and an AR(1) model may well be appropriate. And for nonstationary circumstances randomization is no longer justified in the sense used in strategy A: the fertility of a plot can not be seen as a drawing from a fixed distribution; the variances of the disturbance term differ for different plots (for finite samples this may be evaluated) and the estimate of σ_u^2 loses its meaning. Model-based evaluation of differences in yields remains valid. And the alternating design is far superior to any other design.

The model mentioned above is

$$\begin{aligned} y_i &= X\alpha + d_i + u_i \\ d_i &= d_{i-1} + \xi_i \\ u_i &= \rho u_{i-1} + \epsilon_i \end{aligned} \qquad (12)$$

with ξ_i and ϵ_i independent white noise processes.

This implies

$$(1-L)(1-\rho L)y_i = X^*\alpha + \epsilon_i - \epsilon_{i-1} + \xi_i - \rho\xi_{i-1} = X^*\alpha + (1-\vartheta L)\nu_i \qquad (13)$$

where L is the lag-operator, X^* is a transformation of the matrix X; ν_i is a new white noise process defined by the equivalence $\equiv$, which is based on the equality of variance and covariances of the processes at both sides. These equalities permit one to express ϑ and σ_ν^2 as functions of ρ, σ_ξ^2 and σ_ϵ^2. If σ_ξ^2 is small ν_i resembles ϵ_i, ϑ is close to 1 and the difference with model (2) will be hard to detect empirically.

In the alternating design the first differences contain the information on the difference in yields straightforwardly:

$$(1-L)y_i = (-1)^i\delta + (1-\rho L)^{-1}(1 - \vartheta L)\nu_i \tag{14}$$

with $\delta = \alpha_1 - \alpha_2$

it follows that

$$(1-\rho L)(1-\vartheta L)^{-1}(1-L)y_i = (1+\rho)(1+\vartheta)^{-1}\delta + \nu_i \tag{15}$$

(the proof is simply a matter of writing out the ratio of polynomials in the lag operator and adding the alternating coefficient from (14))
From this one may conclude

$$\mathrm{Var}(d)_{GLS} = n^{-1}(1+\vartheta)^2(1+\rho)^{-2}\sigma_\nu^2 \tag{16}$$

One of the nice features of this formula is the continuous way model (12) converges to model (2) for $\sigma_\xi^2 \Rightarrow 0$. In this case (the nonstationary component becoming less important) $\vartheta \Rightarrow 1$ and $\sigma_\nu^2 \Rightarrow \sigma_\epsilon^2$. And one sees that (16) converges to (9).
So, while in the randomization setup large problems arise in the presence of nonstationary components, the model-based alternating designs are simply generalizations of the stationary case. Also it is clear that approximation of a nonstationary model by a stationary one or vice versa does not lead to severely biased results.

Conclusion

My prejudices against randomizing are strongly confirmed by the exercises above. Randomizing to avoid model dependency is inefficient (there are better designs), leads only on the average to unbiased results and this only in stationary situations. A plausible family of models for correlation structures seems a paradigm to replace that of randomization. The emerging experimental design is robust, as is the resulting inference. The efficiency gains may be large.

Epilogue: the transfer of insights to Monetary Unit Sampling.

Having dealt with old doubts the question arises whether these insights are of any use in a crusade against monetary unit sampling in auditing. In de Vos (1987) a number of arguments are raised and worked out. In this epilogue I try to summarize the main points connected with this article.

First, MUS is a concept where the *conservatism of orthodox sampling methodology* takes on huge proportions. A simple example can make this clear.

Suppose that errors in items of financial statements are distributed Poisson $(n_i\mu)$, with n_i the size of item i and errors independent among items. If this model would be correct, inference on μ could be based on the fact that the total error amount in the sample is distributed Poisson $(\Sigma_S n_i\mu)$. If MUS is applied as if one in every thousand dollars was drawn, exactly the same sample would be evaluated in terms of the number of wrong dollars, being distributed Poisson $(\Sigma_S n_i\mu/1000)$. In other words: 99.9% of the information is ignored.

Unfortunately the model is clearly wrong: in audit populations most items are correct and some are completely or largely wrong. This is in contradiction with the Poisson hypothesis. Cox and Snell (1979) suggested a model with a small probability of an error and an exponential distribution of the error size in case of an error. I.a. Neter and Godfrey (1985) and Moors and Janssens(1987) study the applicability on audit populations. A funny aspect of the Cox-Snell

model is that if it is right small items are as informative as large items, so it is no longer necessary to draw items with probabilities proportional to their sizes (the modern variant of MUS). If it is cheaper to check small items, the auditor could even restrict himself to these. But again, the model is probably wrong. I do not believe that an error of 5 in an item of 10 is as informative as an error of 5000 in an item of 10000.
It would be nice if a robust general model could be developed for audit populations, like in the soil fertility case. This is not simple. Moreover the small number of items containing errors in audit samples prevents efficient estimation of crucial parameters regarding the error size. Empirical Bayesian methods, investigating many different samples to obtain priors for the error size may solve this problem.
As to randomization: apart from the necessity to prevent fraud by giving each item a probability of being checked, there are little arguments for it. In fact if one is interested in the probability of an error conditional upon the reported item, one may develop a systematic design for reported items. Looking at the conditional probabilities has moreover an advantage: the sizes of the reported items that are not in the sample are known, and this information may be used for statements on the total error.

The harm that may be done by orthodox randomization is the main correspondence between the fertility and the audit example. The audit example shows however that the way to a model that can be trusted is not always as easy as the fertility example might suggest.

Appendix: Proof of the equality of OLS and GLS in the alternating design

In the alternating design (4) simplifies to

$$y_i - \rho y_{i-1} = d_{1i}\alpha_1 + d_{2i}\alpha_2 + \epsilon_i \tag{4a}$$

with $d_{1i} = 1$ if i is even and $-\rho$ if i is odd

and $d_{2i} = 1$ if i is odd and $-\rho$ if i is even

The regression formula becomes

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1+\rho^2 & -2\rho \\ -2\rho & 1+\rho^2 \end{bmatrix}^{-1} \begin{bmatrix} (1+\rho^2)b_1 - 2\rho b_2 \\ (1+\rho^2)b_2 - 2\rho b_1 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

with b_1 and b_2 the simple averages for breeds 1 and 2. The variance of $a_1 - a_2$, and thus also the variance of $b_1 - b_2$ is

$$\mathrm{Var}(a_1 - a_2) = (2/n)\sigma_{\epsilon}{}^2{}_* \, (1-\rho^2)^{-2}(2+2\rho^2-4\rho) = (4/n)\sigma_{\epsilon}{}^2{}_*(1+\rho)^{-2}$$

as in (9).

References

Box,G.E.P. and G.C. Tiao (1973) *Bayesian Inference in Statistical Analysis*. Addison-Wesley

Cheng, C. and K. Li(1987) "Optimality criteria in Survey Sampling" *Biometrica*, 74, pp 337-45.

Cox, D.R. and E.J.Snell(1979)"On sampling and the estimation of rare errors" *Biometrica* 66, pp 125-132.

Fienberg, S.E. and J.M. Tanur (1987) "Experimental and Sampling Structures: Parallels Diverging and Meetings" *International Statistical Review*, 55 pp 75-96

Fisher, R.A. (1935) *The Design of experiments*. Edinburgh: Oliver & Boyd.

Holschuh, N.(1980) "Randomization an design: I" in: S.E. Fienberg and D.V. Hinkley eds: *R.A. Fisher: an appreciation*. Springer-Verlag.

Fisher, R.A.(1925) *"Statistical Methods for Research Workers"*. Oliver and Boyd.

Leamer, E.E.(1983) Let's take the con out of Econometrics. American Economic Review

Lopes, L.L. (1982) Doing the impossible: a note on induction and the experience of randomness. *Journal of Experimental Psychology* Vol 8, pp 626-636.

Moors,J.J.A and M.J.B.T. Janssens(1987): "Analytical Properties of Bayesian Cox-Snell bounds in Auditing". Working paper, Catholic University Brabant, Tilburg.

Neter,J. and J.Godfrey(1985) "Robust Bayesian bounds for monetary unit sampling in auditing". *Applied Statistics* 34, pp 157-168.

Picard, R."Randomization an design: I" in: S.E. Fienberg and D.V. Hinkley eds: *R.A. Fisher: an appreciation*. Springer-Verlag.

Sneek, J.M.(1985) *"Modeling procedures for univariate economic time series"* Free University Press, Amsterdam.

"Student"(1936). "Co-operation in Large Scale Experiments" *Supplement to the Journal of the Royal Statistical Society*, 4, pp 115-136.

"Student"(1937). "Comparison Between Balanced and Random Arrangements of Field Plots," *Biometrica*, 29, 363-136.

Vos, A.F. de(1987) *"Bayesian and Empirical Bayesian Inference with Audit Applications"* Preliminary version of monograph, obtainable from the author.

1986-48	I. Evers, M. Fischer P. Nijkamp	A cross-national comparative analysis of regional labour markets
1986-49	E.J. Davelaar P. Nijkamp	Spatial dispersion of technological innovation: the incubator hypothesis
1986-50	W. Barentsen P. Nijkamp	Modelling non-linear processes in time and space
1986-51	P. Nijkamp A. Reggiani	Analysis of dynamic spatial interaction models by means of optimal control
1986-52	J. de Groot	Dominicaanse republiek Landhervorming in de suikersector
1986-53	H. Clemens	Modernisering van de landbouw in socialistische perifere economieën
1986-54	A.J. Vermaat	Groepsvorming bij rationeel gedrag

1987-1	P. Rietveld	On Multidimensional Inequality Comparisons
1987-2	H.J. Bierens	A consistent Hansman-type Model Specification Test
1987-3	H. Visser	A Survey of Recent Developments in Monetary Theory
1987-4	E. Eeftink D. Korf	Externe verslaggeving van beleggingsinstellingen
1987-5	Pitou van Dijck	Transforming the trade and industrialization regime in developing countries
1987-6	Pitou van Dijck	The strong factor-intensity assumption reconsidered
1987-7	P. van Dijck en H. Verbruggen	The gains from trade for developing countries reconsidered
1987-8	H. Visser	Macroeconomische aspecten van bedrijfs-economisch toezicht
1987-9	F. van der Wel Th. de Wit	Stelselwijzigingen in de jaarrekening: verslag van een empirisch onderzoek
1987-10	J. de Groot H. Clemens	Agrarian labour market and technology under different regimes: A comparison of cuba and the Dominican Republic
1987-11	I.J. Steyn A.F. de Vos	Structural time series models for trends
1987-12	J.P. de Groot	Collective Rice-Farms in the Dominican Republic
1987-13	R.W. Veldhuizen	Valuta Management en Management Control
1987-14	J. Koelewijn	De achtergronden van het verdwijnen van de zelfstandige hypotheekbanken in de jaren tachtig
1987-15	H.C. Tijms	A quick and practical apporximation to the waiting time distribution in the multi-server queue with priorities
1987-16	H.C. Tijms	Educatieve Operations Research Software: Wis en Waarachtig
1987-17	F.C. Palm and C.C.A. Winder	The life cycle consumption model under structural changes in income and moving planning horizons
1987-18	H.J. Bierens	Basic probability theory
1987-19	H.J. Bierens	Convergence
1987-20	H.J. Bierens	Introduction to conditioning
1987-21	H.J. Bierens	Nonlinear parametric regression analysis
1987-22	H.J. Bierens	Tests for model misspecification
1987-23	T. van der Meer	New perspective on price indices
1987-24	S. Flejterski	Theoretische Aspecten van de Exportdiversificatie
1987-25	J. Rouwendal	A note on discrete choice under uncertainty
1987-26	P.H.M. Ruys en G. van der Laan	Computation of an industrial equilibrium
1987-27	H.G. Eijgenhuijsen J. Koelewijn H. Visser	Groeibelemmerende factoren en de rol van financiële instellingen bij de financiering van investeringen
1987-28	J.C.W. van Ommeren	Asymptotically exponential expansion for the waiting time probability in the single server queue with batch arrivals
1987-29	R.D. Nobel	Practical approximations for finite-buffer queueing models with batch-arrivals
1987-30	H. Linnemann C. van Beers	Measures of export-import similarity, and the linder hypothesis once again
1987-31	W. van Lierop H. de Neef	Dynamic Analyses with loglinear and disaggregate choice models